Minimización Cuasi-convexa: teoría, métodos y aplicaciones

Erik Alex Papa Quiroz
Facultad de Ciencias Matemáticas
Universidad Nacional Mayor de San Marcos, Perú
COPIOS 2019

UNSA-UNMSM
Arequipa-Perú

24/10/2019

Resumen

1. Motivación

- Computer vision (Visión Computacional): Quasiconvex optimization for robust geometric reconstruction, Ke and Kanade (2007)
- Spanning economics:
  -Price discrimination and social Welfare, Varian (1985): marginal cost, market prices, utility functions, etc.
  -The theory of incentives: the principal agent model, Laffont and Martimort (2009): Book

- Industrial organization: Topics in microeconomics: Industrial organization, auctions and incentives, Wolfstetter (1999)
- Location Theory: Quasiconvex optimization and location theory. Gromicho (1998)
- Applied Mathematics: Nash Equilibrium Problems, quasiconvex multicriteria minimization, fractional optimization, between others.

Ejemplo: Teoría Económica de la Decisión

Los modelos en la teoría económica de la decisión consiste en la mejor elección del agente económico dentro de todas las posibles alternativas a escoger.

- El conjunto de elección, que nos dice cual es el universo de alternativas.
- El criterio de valorización se define mediante una relación binaria $\precsim$ (preferencia)
- Las restricciones, que delimitan el conjunto de oportunidades.

Modelo

Dado el conjunto de elección X , una función $u : X \to R$ es una función de utilidad para $\precsim$ si para todo $x, y \in X$

$$y \precsim x \Leftrightarrow \mu\left(y\right) \leq \mu\left(x\right).$$

Las preferencias $\precsim$ bajo ciertas condiciones sobre X pueden representarse por una función de utilidad μ . Elegir la mejor alternativa se convierte así en

$$\max\{\mu(x) : x \in X\}$$

Un tipo particular de función de utilidad es la función cuasi-cóncava que está íntimamente relacionada a la hipótesis de convexidad de la preferencia. Recordemos que $\precsim$ es convexa si dados $x, y \in X$ con $x \precsim y$ y $x \precsim z$ en X y $0 \leq \alpha \leq 1$ entonces

$$x \precsim \alpha y + (1 - \alpha) z.$$

Recordemos un resultado bien conocido:
$\precsim$ es convexa si, y solamente si, μ es cuasi-cóncava. Asi el problema de optimización:

$$\max\{\mu(x) : x \in X\}$$

es cuasicóncavo.

Ejemplos de funciones cuasi-cóncavas en economia son:

- La función de Cobb-Douglas $\mu : \mathbb{R}^n_+ \to \mathbb{R}$ such that

$$\mu(x) = k \prod_{i=1}^{n} x_i^{\alpha_i},$$

  donde $\alpha_i > 0, \forall i, \sum_{i=1}^{n} \alpha_i > 1,$ y $k > 0.$

- La función de producción C.E.S $\mu : \mathbb{R}^n_+ \mathrm{ri} \mathbb{R},$ tal que

$$\mu(x) = k \left[\sum_{i=1}^{n} \delta_i x_i^{-\rho} \right]^{-v/\rho}$$

  donde $\delta_i \in (0,1), \forall i, \sum_{i=1}^{n} \delta_i = 1, k > 0, v \in (0,1)$ y $\rho > -1, \rho \neq 0.$

2. Teoría

Definición: Función convexa

Sea $f : C \subset \mathbb{R}^n \to \mathbb{R}$, donde C es un conjunto convexo y $C \neq \emptyset$. La función f es llamada convexa si

$$f(\alpha x_1 + (1-\alpha)x_2) \leq \alpha f(x_1) + (1-\alpha)f(x_2)$$

para todo $x_1, x_2 \in C$ y $\alpha \in [0, 1]$.
f será estrictamente convexa en C si la desigualdad es estricta para todo $x_1, x_2 \in C$, $x_1 \neq x_2$, y $\alpha \in (0, 1)$.

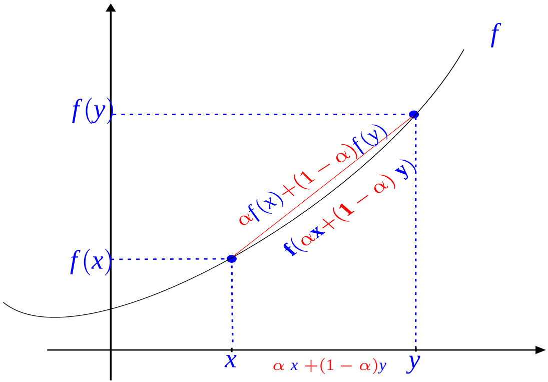

Figure: Gráfica de una función convexa

Definición

Sea $f : C \subset R^n \to R$ una función definida en el conjunto convexo C. Entonces f es cuasi-convexa si

$$f\left(\alpha x + (1-\alpha) y\right) \leq \max\left\{f(x), f(y)\right\}$$

$\forall\, x, y \in C, \forall\, \lambda \in [0, 1]$.

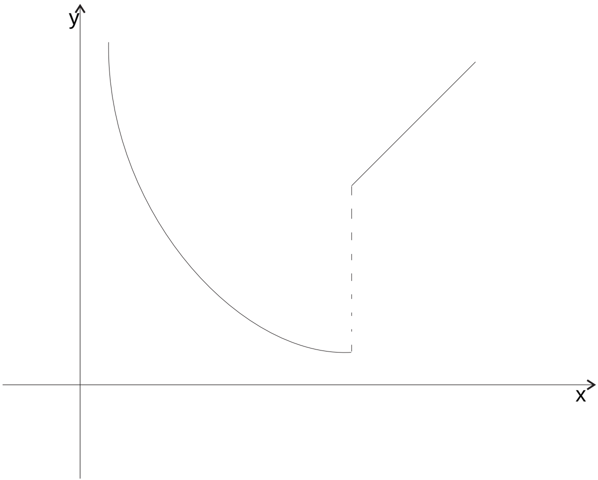

Figure: Función cuasi-convexa discontinua

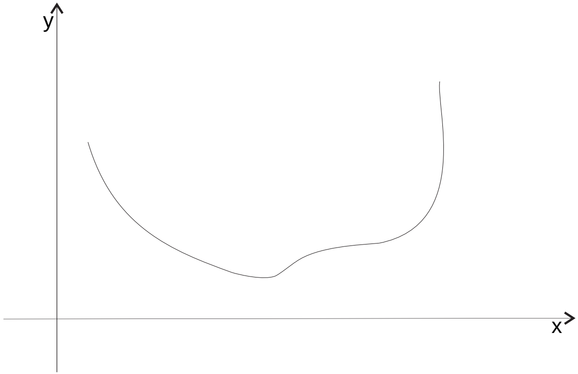

Figure: Función cuasi-convexa continua

Caracterizaciones y propiedades

- f es cuasi-conexa si sus conjuntos de nivel inferior

$$L(f, \alpha) = \{x \in C : f(x) \leq \alpha\}$$

son conjuntos convexos para todo $\alpha \in R$

- (Fenchel, 1951) Sea $\phi : C \subset R^n \rightarrow R$ una función cuasi-convexa y sea $f : D \subset R \rightarrow R$ una función no decreciente en D conteniendo la imagen de ϕ, entonces la función composición $f \circ \phi$ es también cuasi-convexa.

Diferencias con funciones convexas

- Cuando la función es convexa es bien conocido que todo mínimo local es global. Sin embargo, para funciones cuasi-convexas este resultado no es verdadero. Basta considerar la función cuasi-convexa.

$$f(x) = \begin{cases} x^2 & -1 \leq x \leq 1 \\ 2 & 1 < x \leq 3 \\ -(x-4)^2 + 6 & 3 < x \leq 4. \end{cases}$$

Observamos que $\bar{x} = 2$ es un mínimo local, pero no un mínimo global.

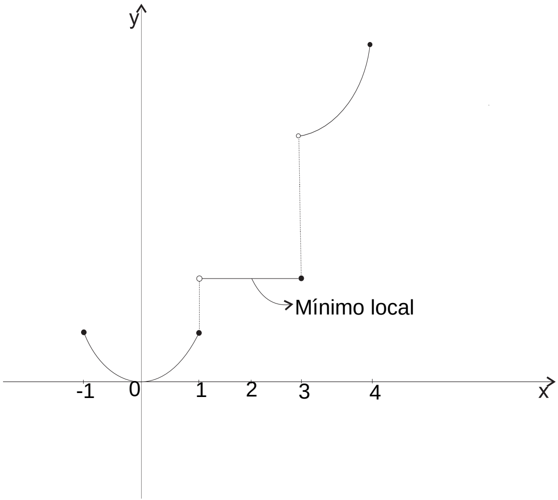

Figure: Mínimo local, que no es mínimo global

Funciones cuasi-convexas diferenciables

- Sea C un conjunto convexo abierto de $\mathbb{R}^n$ y $f : C\mathrm{ri}\mathbb{R}$ una función diferenciable en C. Entonces, f es convexa si y solo si

$$f(y) \geq f(x) + \langle \nabla f(x), y - x \rangle, \forall x, y \in C.$$

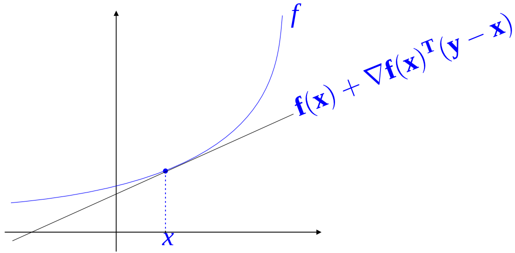

Figure: Gráfica de una función convexa

- (Arrow y Enthoven, 1961) Sea $f : C \subset R^n \to R$ una función dife-renciable en el conjunto abierto convexo C. Entonces, f es una función cuasi-convexa si y solo si para todo $x_1, x_2 \in C$

$$f(x_1) \leq f(x_2) \;\; implica \;\; que \;\; (x_1 - x_2)^T \nabla f(x_2) \leq 0$$

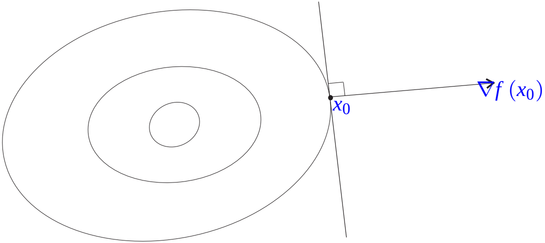

Figure: Conjuntos de nivel de una función cuasi-convexa

Funciones cuasi-convexas dos veces diferenciables

- Sea C un conjunto convexo abierto de $\mathbb{R}^n$ y $f : C\mathrm{ri}\mathbb{R}$ una función dos veces continuamente diferenciable en C. Entonces, f es convexa si y solo si

$$\nabla^2 f(x) \succeq 0, \forall x \in C.$$

- (Arrow y Enthoven, 1961; Avriel, 1972) Sea una función $f : C \subset R^n \to R$ cuasi-convexa y dos veces continuamente diferenciable en el conjunto abierto y convexo C. Si $x_0 \in C$, $v \in R^n$ y $v^T \nabla f\left(x_0\right) = 0$ entonces $0 \leq v^T \nabla^2 f\left(x_0\right) v$.

- Sea $f : C \subset R^n \to R$ una función cuasi-convexa dos veces continua-mente diferenciable en el conjunto abierto C. Si $x_0 \in C$ y $\nabla f(x_0) = 0$, entonces $0 \leq z^T \nabla^2 f(x_0) z$ para todo $z \in R^n$.
- (Gerencsér, 1973) Sea $f : C \subset R^n \to R$ una función cuasi-convexa dos veces continuamente diferenciable en el conjunto abierto y convexo C. Entonces $\nabla^2 f$, la Hesiana de f, tiene a lo más un autovalor negativo para todo $x \in C$.

Condición necesaria y suficiente para funciones cuasi-convexas dos veces diferenciables

El k-ésimo orden de la matriz Hessiana de una función f dos veces continuamente diferenciable en el punto $x \in R^n$ es definida como

$$D_k(x) = \begin{vmatrix} 0 & \frac{\partial f}{\partial x_1} & \frac{\partial f}{\partial x_2} \cdots & \frac{\partial f}{\partial x_k} \\ \frac{\partial f}{\partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_1} & \frac{\partial^2 f}{\partial x_1 \partial x_2} \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_k} \\ \cdots & \cdots & & \cdots \\ \frac{\partial f}{\partial x_k} & \frac{\partial^2 f}{\partial x_k \partial x_1} & \frac{\partial^2 f}{\partial x_k \partial x_2} \cdots & \frac{\partial^2 f}{\partial x_k \partial x_k} \end{vmatrix}.$$

- Una condición necesaria para que f sea cuasi-convexa en un conjunto convexo $C \subset R^n$ es

$$(-1)^k \det D_k(x) \leq 0, \quad k = 1, ..., n$$

para todo $x \in C$, donde det denota la determinante.
- Una condición suficiente para una función $f : C \subset R^n \to R$ dos veces continuamente diferenciable en un conjunto convexo C sea cuasi-convexa en C es que

$$(-1)^k \det D_k(x) < 0, \quad k = 1, ..., n$$

para todo $x \in C$.

Ejemplo

La función

$$f(x_1, x_2, x_3) = x_1 x_2 x_3$$

es cuasi-cóncava en R^3_{++}.

En efecto,

$$D_3(x) = \begin{vmatrix} 0 & x_2x_3 & x_1x_3 & x_1x_2 \\ x_2x_3 & 0 & x_3 & x_2 \\ x_1x_3 & x_3 & 0 & x_1 \\ x_1x_2 & x_2 & x_1 & 0 \end{vmatrix}$$

Tenemos

$$|D_1| = -x_2^2 x_3^2 < 0$$

$$|D_2| = 2x_1 x_2 x_3^3 > 0$$

$$|D_3| = -x_1^2 x_2^2 x_3^2 < 0$$

entonces

$$(-1)^1 |D_1| > 0$$

$$(-1)^2 |D_2| > 0$$

$$(-1)^3 |D_3| > 0$$

Por lo tanto, $f(x_1, x_2, x_3) = x_1 x_2 x_3$ es cuasi-cóncava en R_{++}^3.

3. Métodos para optimización cuasi-conevxa

El modelo es:

$$\begin{cases} \min \ f(x) \\ s.a \\ x \in X \subset \mathbb{R}^n \end{cases}$$

donde $f : X \subset \mathbb{R}^n \to \mathbb{R}$ es una función cuasi-convexa, i.e.,

$$f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}, \forall x, y \in H; \forall \lambda \in [0,1]$$

y X es un conjunto convexo.

- Método del gradiente (Kiwiel and Murty 1996)
- Método subgradiente (Kiwiel 2001)
- Métodos de los multiplicadores (??)
- Métodos proximales (Papa Quiroz and Oliveira 2009)
- Métodos de penalidades (??)
- Método de Newton (??)
- Método de Quasi-Newton (??)
- Método de Gradiente conjugado (??)
- Métodos proyectivos (??)

4. Método del Punto Proximal

Consideremos el problema

$$\begin{cases} \min \ f(x) \\ s.a \\ \quad x \in H \end{cases}$$

El método del punto proximal genera una sucesión de puntos $\{x^k\}$ tal que:

$$\begin{cases} x^0 \in H \\ x^k = argmin\{f(x) + \frac{\lambda_k}{2}\|x - x^{k-1}\|^2, x \in H\} \end{cases}$$

Este método fue introducido por Martinet (1970) para resolver el problema cuando $f : H \to \mathbb{R} \cup \{\pm\infty\}$ es una función propia, convexa, semicontínua inferior y H es un espacio de Hilbert.

Fue demostrado que si f es propia, convexa, semicontinua inferior y si la sucesión $\{\lambda_k\}$ satisface:

$$\sum_{k=1}^{+\infty} \frac{1}{\lambda_k} = +\infty.$$

entonces $\{f(x^k)\}$ converge al ínfimo de f y si además el conjunto de soluciones óptimas no es vacio, entonces $\{x^k\}$ converge débil (fuerte si la dimensión es finita) a una solución del problema, ver Guler (1991).

4.1 Importancia

- Se ha convertido en la base teórica para justificar la convergencia de los métodos de multiplicadores para resolver problemas de optimización. En particular, se ha demostrado que detrás de cada método multiplicador existe un método proximal que lo genera (Rockafellar, 1976).

- Se ha probado también que una gran clase de métodos de descomposición de problemas convexos son casos particulares del método del punto proximal para encontrar un cero de un operador monótono maximal, específicamente el método de Douglas-Rachford estudiado por Lions y Mercier (que abarca una gran clase de métodos de optimización) es una versión particular del método proximal, (Eckstein y Bertsekas, 1992).

4.2 Algunos trabajos relacionados

- Tseng (JOTA, 2001), probó un resultado de convergencia débil, i.e, si la función objetivo es semicontinua inferior y acotada inferiormente y además $\lambda_k = \lambda > 0$ entonces todo punto de acumulación z es un punto estacionario:

$$f'(z,d) := \liminf_{\lambda \downarrow 0} \left(\frac{f(z+\lambda d) - f(z)}{\lambda} \right) \geq 0.$$

- Kaplan y Tichatschke (JOGO, 1998), estudió el método para una clase de funciones no convexas cuando la función auxiliar $f(.) + (\lambda_k/2)||. - x^{k-1}||^2$ se convierte en fuertemente convexa sobre cierto conjunto convexo mediante la elección de un cierto λ_k.

4.2 Algunos trabajos relacionados (continuación)

- Attouch y Bolte (Math Prog, 2009), bajo la hipótesis que f satisface una propiedad de Lojasiewicz y $\{x^k\}$ acotada, probaron la convergencia del método para algún punto crítico generalizado.
- Attouch y Teboulle (JOTA, 2004) y Alvarez et al (SIAM J.Optim, 2004), para f cuasi-convexa y con un sistema continuo prueban la convergencia del método para un cierto conjunto que contiene al conjunto de soluciones óptimas.

4.2 Algunos trabajos relacionados (continuación)

- Cunha et al.(Optimization, 2010) y Chen and Pan (Pacific J.Optim, 2008): para f cuasi-convexa diferenciable y usando distancias $\phi-$divergencias particulares prueban la convergencia a un punto KKT cuando el parámetro es acotado.
- Pan y Chen (JOGO, 2007): para f cuasi-convexa no necesariamente diferenciable y usando distancias homogéneas de segundo orden prueban la convergencia del método a un punto KKT cuando el parámetro es acotado.
- Souza et al (EJOR, 2010). para f cuasi-convexa diferenciable y usando distancias separables de Bregman prueban la convergencia del método a un punto KKT cuando el parámetro es acotado.

5. Extensión del Método

Estamos interesados en resolver el problema

$$\min\{f(x) : x \geq 0\}$$

donde $f: \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ es una función propia, semicontinua inferior, localmente Lipschitz y cuasi-convexa tal que $domf \cap \mathbb{R}^n_+ \neq \phi$ y $x \geq 0$ significa que cada componente de x, x_i, es no negativo.

5.1 Algoritmo

Dados una secuencia de parámetros positivos $\{\lambda_k\}$ y un punto inicial

$$x^0 \in \mathbb{R}^n_{++}. \tag{1}$$

Para cada $k = 1, 2, \ldots,$ si $0 \in \widehat{\partial} f(x^{k-1}),$ entonces parar. caso contrario, encontrar $x^k \in \mathbb{R}^n$ tal que

$$0 \in \widehat{\partial}(f(\cdot) + \lambda_k d(\cdot, x^{k-1}))(x^k). \tag{2}$$

donde $\widehat{\partial}$ es el subdiferencial de Clarke y d es una distancia proximal tal que $(d, H) \in F_+(\mathbb{R}^n_+).$

5.2 Hipótesis

1. **Hipótesis A.** $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ es una función propia acotada inferiormente y localmente Lipschitz.
2. **Hipótesis B.** $f : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ es una función semicontínua inferior y cuasi-convexa.

Como estamos interesados en la convergencia asintótica del método, también asumiremos que en cada iteración $0 \notin \widehat{\partial} f(x^k)$ esto implica que $x^k \neq x^{k-1}$, para todo k.

5.3 Algunas Propiedades

- Bajo las hipótesis dadas se tiene que $\{f(x^k)\}$ es una secuencia decreciente y además convergente.

Definamos el siguiente conjunto

$$U_+ := \{x \in dom\, f \cap \mathbb{R}^n_+ : f(x) \leq \inf_{j \in \mathbb{N}} f(x^j)\}.$$

Observemos que este conjunto depende de la elección de la iteración inicial x^0 y la secuencia $\{\lambda_k\}$.
Si $U = \emptyset$ entonces

i) $\lim_{k \to +\infty} f(x^k) = \inf_{x \geq 0} f(x)$,

ii) $\{x^k\}$ es no acotada.

5.4 Convergencia

- Bajo las hipótesis A, B y $(d, H) \in \mathcal{F}_+(\mathbb{R}^n_+)$,, la secuencia $\{x^k\}$ converge a algún punto de U_+.
- Si f es continuamente diferenciable, cuasi-convexa, acotada inferiormente y $0 < \lambda_k < \bar{\lambda}$, entonces la secuencia $\{x^k\}$ converge a un punto KKT del problema. Además, si el problema tiene solución y existe $i \in I(\bar{x})$ tal que

$$\limsup_{k \to +\infty} \left(\lambda_k \frac{\partial d_i}{\partial x_i}(x_i^k, x_i^{k-1}) \right) < 0,$$

entonces $\{x^k\}$ converge a una solución óptima del problema.

5.5 Otros Resultados

- Si el subdiferencial es de Frêchet (no se asume ahora que f es localmente Lipschitz) entonces se obtiene los mismos resultados de convergencia (Papa Quiroz y Oliveira, EJOR 2012).
- Si la restricción del problema es $C = \{x : Ax \leq b\}$, usando el subdiferencial de Clarke y distancias de Bregman se obtiene la convergencia de la sucesión a un punto crítico del problema (Papa quiroz y Mallma Raimirez, 2012).
- Si el conjunto de restricciones está definido sobre una variedad de Hadamard, usando distancias de Bregman y subdiferencial de Frêchet sobre la variedad, la sucesión converge a cierto conjunto que incluye el conjunto de soluciones óptimas (Papa Quiroz, 2012).

6. Referencias

ABB04 Alvarez, F., Bolte, J., and Brahic, O., Hessian Riemannian Gradient Flows in Convex Programming, SIAM J. Optim., 43, 2, 477-501, 2004.

Attouch, H., and Bolte, J. On the Convergence of the Proximal Algorithm for Nonsmooth Functions Involving Analytic Features, Mathematical Programming, 116, 1-2, B, 5-16, 2009.

Attouch, H., and Teboulle, M. Regularized Lotka-Volterra Dynamical System as Continuous Proximal-Like Method in Optimization, Journal Optimization Theory and Applications, 121, 541-580, 2004.

Chen, J. S., and S. Pan, S. H. Proximal-Like Algorithm for a Class of Nonconvex Programming, Pacific Journal of Optimization, 4, 319-333, 2008.

Cunha, G. F. M., da Cruz Neto, J. X., and Oliveira, P. R. A Proximal Point Algorithm with ϕ-Divergence to Quasiconvex Programming, Optimization, 59, 777-792, 2010.

Eckstein J. y Bertsekas D.P. On the Douglas-Rachford splitting method and the proximal algorithm for maximal monotone operators. Mathematical Programming Vol 55, N°3: 293-318, 1992.

Gromicho J. Quasiconvex optimization and location theory. Kluwer Academic Publishers, Dordrecht, the Netherlands, (1998).

Guler O. On the Convergence of the Proximal Point Algorithm for Convex Minimization. SIAM J. Control and Optimization, 29 (2), 403-4019, 1991.

Kaplan, A., and Tichatschke, R. Proximal Point Methods and Nonconvex Optimization, Journal of global Optimization, 13, 389-406, 1998.

Martinet B. Régularisaton, D'inéquations Variationelles par Approximations Successives, Revue Française D'Informatique et de Recherche Operationelle, 154-159, 1970.

Mas-Colell A, Whinston MD, Green JR. Microeconomic theory, Oxford University Press: New York, NY, USA; 1995.

Pan, S. H., and Chen, J. S. Entropy-Like Proximal Algorithms Based on a Second-Order Homogeneus Distances Function for Quasiconvex Programming, Journal of Global Optimization, 39, 555-575, 2007.

Papa Quiroz, E.A., Oliveira, P.R. Proximal Point Methods for Quasiconvex and Convex Functions with Bregman Distances on Hadamard Manifolds. Journal of Convex Analysis, 16,1, pp 49-69, 2009.

Papa Quiroz, E.A., Oliveira, P.R. Full Convergence of the Proximal Point Method for Quasiconvex Functions on Hadamard Manifolds. ESAIM: Control, Optimization and Calculus of Variations, v.18(2), pp. 483-500, 2012.

Papa Quiroz, E.A. An Extension of the Proximal Point Algorithm with Bregman Distances on Hadamard Manifolds, Sometido a JOGO, 2011.

Papa Quiroz, E.A., Mallma Ramirez, L. Proximal Point Algorithm with Generalized Distances for Quasiconvex functions with linear inequalities, Sometido a EJOR, 2012.

Papa Quiroz, E.A., Oliveira, P.R. Proximal Point Method for Minimizing Quasiconvex Locally Lipschitz Functions on Hadamard Manifolds, Nonlinear Analysis, 75, pp. 5924-5932, 2012.

Papa Quiroz, E.A., Oliveira, P.R. An extension of proximal methods for quasiconvex minimization on the nonnegative orthant, European Journal of Operational Research, 2016, pp. 26-32, 2012.

Rockafellar, R. T. *Augmented Lagrangians and Application of Proximal Point Algorithm in Convex Programming*, Mathematics of Operation Research, Vol.1, pp. 97-116, 1976.

Takayama A. Mathematical economics, 2nd Edition, Cambridge Univ. Press, 1995.

Tseng P. Convergence of a Block Coordinate Descent Method for Nondifferentiable Minimization, Journal of Optimization Theory and Applications, 109, 3, 475-494, 2001.

Souza, S. S., Oliveira, P. R., da Cruz Neto, J. X., and Soubeyran, A. A Proximal Method with Separable Bregman Distance for Quasiconvex Minimization on the Nonnegative Orthant, European Journal of Operational Research, 201, 365-376, 2010.